Phase retrieval with the transport-of-intensity equation: matrix solution with use of Zernike polynomials

T. E. Gureyev, A. Roberts, and K. A. Nugent

School of Physics, The University of Melbourne, Parkville, Victoria 3052, Australia

Received October 3, 1994; revised manuscript received January 11, 1995; accepted March 29, 1995

A new technique is proposed for the recovery of optical phase from intensity information. The method is based on the decomposition of the transport-of-intensity equation into a series of Zernike polynomials. An explicit matrix formula is derived, expressing the Zernike coefficients of the phase as functions of the Zernike coefficients of the wave-front curvature inside the aperture and the Fourier coefficients of the wave-front boundary slopes. Analytical expressions are given, as well as a numerical example of the corresponding phase retrieval matrix. This work lays the basis for an effective algorithm for fast and accurate phase retrieval.

1. INTRODUCTION

The problem of phase retrieval from intensity information is of relevance in many areas of science.[1,2] In this paper we consider the phase retrieval technique based on the transport-of-intensity equation (TIE) first proposed by Teague.[3,4] We are motivated to investigate this problem as the TIE has been studied in the context of adaptive optics for astronomy[5,6] and the imaging of phase objects in microscopy.[7] Related problems have also benefited from the sorts of approach described here, such as determination of the aberrations of the eye in ophthalmology,[8] correcting optics for x-ray sources,[9] and the investigation of aberrations in large optical telescopes.[10]

In this paper we are concerned not with the nature of a specific application but rather with the detailed mathematical basis for the solution of the TIE. We concentrate here on solutions in terms of the Zernike polynomials,[11] as these are the natural starting point for the discussion of the diffraction theory of aberrations and so are widely used in adaptive optics and other optical studies.[12-19] Although the details of our discussion revolve around these polynomials, it is no doubt also possible to develop closely related approaches with the use of other orthonormal polynomials, which may be useful in specific applications.

The TIE is a partial differential equation that directly relates the phase distribution in the planes orthogonal to the optical axis to the rate of change of the wave-front intensity of the beam. The equation forms the basis of the now widely used wave-front curvature sensing technique proposed by Roddier and Roddier.[5,6,10] Here we suggest a new approach based on Zernike decomposition of the TIE that we hope will allow phase to be sensed with improved speed and resolution. We propose to expand each function involved in the TIE into a series of Zernike polynomials, thus reducing the boundary-value problem for the TIE to a system of linear algebraic equations. Such an approach is effective in the case of circular apertures with uniform illumination, to which we confine our present study. The analysis of the structure of the resulting algebraic system significantly clarifies the contribution of each particular phase aberration to the propagation of the wave-front intensity distribution in the beam. Furthermore, the reduction of the TIE to a system of linear equations allows simple and efficient methods for its solution. Some facts about the algebraic properties of the TIE with respect to Zernike polynomials have been reported earlier.[10-19] We give rigorous proofs for the relevant results, derive some new ones, and bring them together to present a complete picture of the structure of the TIE with respect to individual Zernike aberrations of the phase.

Our study is based on the calculation of the Laplacian of Zernike polynomials. These results can be considered as a logical extension of the work by Noll,[13] who calculated the first-order partial derivatives of Zernike polynomials. We prove that the only Zernike polynomials with zero Laplacian are the diagonal ones, i.e., those with radial degree equal to the azimuthal frequency. We also prove that the Laplacian of any nondiagonal Zernike polynomial of radial degree N can be represented as a linear combination of Zernike polynomials each with radial degree not exceeding $N - 2$. Using these results, we derive an explicit operator formula for the phase solution to the TIE as a function of Zernike coefficients of the wave-front curvature inside the aperture and the Fourier coefficients of the wave-front slopes at the boundary. It turns out that, on account of the stability of the solution to the Neumann problem for the Poisson equation, this phase solution is insensitive to small errors in the data of the wave-front curvature and boundary slopes. Furthermore, the coefficients of the inverse operator do not depend on the experimental data. Therefore we believe that our approach lays the basis for a very efficient algorithm for phase retrieval by the TIE method.

In Section 2 we review the basics of the phase reconstruction with the TIE. In Section 3 we develop our new approach, and we discuss the results in Section 4.

2. TRANSPORT-OF-INTENSITY EQUATION AND PHASE RETRIEVAL

The underlying idea of phase retrieval with the use of the TIE is that in the paraxial approximation the evolution of the intensity distribution in the direction of beam

0740-3232/95/091932-10$06.00

propagation is defined mainly by the distribution of the phase in the planes orthogonal to that direction. Therefore these phase distributions can be recovered if the intensity change from one such plane to another is measured. In this section we recall the basics of the intensity transport method, as suggested by Teague[3,4] and Roddier and Roddier.[5,6,10]

Let us consider the scalar monochromatic electromagnetic wave with complex amplitude

$$\exp(ikz)u(\mathbf{r}) = I^{1/2}(\mathbf{r})\exp[ikz + i\phi(\mathbf{r})], \tag{1}$$

where $\mathbf{r} = (x, y, z)$. In the paraxial (Fresnel) approximation with the optical axis parallel to z, the complex amplitude $u(\mathbf{r})$ satisfies the paraxial equation

$$(2ik\partial_z + \Delta)u(x, y, z) = 0, \tag{2}$$

where k is the wave number, $\partial_z = \partial/\partial z$, and $\Delta = \nabla^2 = \partial_x{}^2 + \partial_y{}^2$ is the two-dimensional Laplacian. If we substitute Eq. (1) into Eq. (2) and separate real and imaginary parts, we obtain the following pair of equations [provided that $I(\mathbf{r}) \neq 0$]:

$$2k\partial_z\phi = -|\nabla\phi|^2 + D(I), \tag{3}$$

$$k\partial_z I = -\nabla I \cdot \nabla\phi - I\Delta\phi, \tag{4}$$

where $\nabla = (\partial_x, \partial_y)$ is the gradient operator in a plane and $D(I) = I^{-1/2}\Delta(I^{1/2})$ is the diffraction term. Equation (4) is the TIE. It can be used for the reconstruction of the phase in some area Ω of a plane (x, y), $z =$ constant, if the distributions of intensity and its z derivative are known there.

In this work we will consider only circular domains Ω, where R is the radius and Γ is the boundary of Ω. It is convenient to introduce the polar coordinates (r, θ) in the plane of interest, $z = 0$. We also restrict our study to the case of uniform intensity distributions in Ω:

$$I(r, \theta) = I_0 H(R - r), \qquad I_0 = \text{constant},$$
$$H(t) = \begin{cases} 1 & t > 0 \\ 0 & t \leq 0 \end{cases}. \tag{5}$$

The assumption of uniform intensity is widely accepted in adaptive optics.[5,6,20] Substituting Eqs. (5) into Eq. (4), we obtain

$$-H(R - r)\Delta\phi(r, \theta) + \delta(R - r)\partial_r\phi(R, \theta) = kI_0{}^{-1}\partial_z I(r, \theta), \tag{6}$$

where $\delta(r)$ is the Dirac delta function and $\partial_r\phi$ is the phase derivative along the radial direction. Equation (6) implies that the z derivative of intensity must also contain a delta-function term at the boundary:

$$kI_0{}^{-1}\partial_z I(r, \theta) = f(r, \theta) + \delta(R - r)\psi(\theta), \tag{7}$$

where the function f is smooth up to the boundary and ψ is a smooth function on the boundary Γ. Comparing Eqs. (6) and (7), we find that

$$-\Delta\phi = f \tag{8}$$

inside the circular domain Ω and that

$$\partial_r\phi = \psi \tag{9}$$

on the boundary Γ.

Thus, in the case of uniform intensity, the phase can be obtained as a solution to the Neumann boundary-value problem (9) for the Poisson equation (8). This approach was developed by Roddier[5,6] and has an important advantage over the original suggestion of Teague,[3,4] who considered the Dirichlet boundary conditions $\phi = \tilde{\psi}$ on Γ. The advantage of boundary conditions (9) is in the fact that, unlike the phase itself, the phase normal derivative at the boundary can be found from intensity measurements (7) at this boundary. Hence direct measurements of the phase boundary values are not necessary.

When one studies a boundary problem for a partial differential equation, it is always necessary to address three major questions, namely, those concerning existence, uniqueness, and stability of solutions. Because the Neumann problem [Eqs. (8) and (9)] is a classical object of mathematical physics, its properties are well known.

It is proved in the theory of partial differential equations that a solution to the problem of Eqs. (8) and (9) exists if and only if the following condition holds[21]:

$$\iint_\Omega f(r, \theta) r \,\mathrm{d}r \mathrm{d}\theta + \int_\Gamma \psi(\theta) R \,\mathrm{d}\theta = 0. \tag{10}$$

When we substitute for f and ψ, Eq. (10) becomes

$$k\int_0^{2\pi}\int_0^R \partial_z I(r, \theta) r \,\mathrm{d}r\mathrm{d}\theta = -I_0 R\int_0^{2\pi} \partial_r\phi(R, \theta)\mathrm{d}\theta, \tag{11}$$

which is just an expression of conservation of energy; loss of intensity in a region arises through energy flow across the boundary of the region. Equation (10) may be used to check the consistency of acquired intensity data.

The mathematical theory also states[21] that the solution of Eqs. (8) and (9) is unique up to a constant, i.e., if ϕ is a solution, then $\phi + C$ is also a solution for any constant C. This arbitrary additive constant is not essential for the phase reconstruction. A nontrivial fact is that in the case of uniform intensity (5) and circular domain Ω the phase reconstructed by Eqs. (8) and (9) is unique (up to a constant) even in the class of multivalued phase functions.[22] This is important in view of the example given by Gori *et al.*,[23] which presents essentially different (multivalued) phase functions corresponding to the same (nonuniform) three-dimensional intensity distribution in a wave field.

Finally, we would like to recall that the solution ϕ to Eqs. (8) and (9) is stable with respect to small errors in f or ψ, as a result of the boundedness of the inverse operator.[21,24] Namely, if ϕ and ϕ' are the solutions to Eqs. (8) and (9) with the right-hand-side functions (f, ψ) and (f', ψ'), respectively, with $d_\Omega(f, f') < \delta_1$ and $d_\Gamma(\psi, \psi') < \delta_2$, where d_Ω and d_Γ are the appropriate metrics inside Ω and on the boundary Γ, respectively, then $d_\Omega(\phi, \phi') < \epsilon$ and $\epsilon = \epsilon(\delta_1, \delta_2) \to 0$ when $\delta_1 + \delta_2 \to 0$.

Note that, in the approach described above, the boundary values of the phase normal derivative $\psi(\theta) = \partial_r\phi(R, \theta)$ should be obtained as the coefficient of the delta function in Eq. (7). In reality, the intensity change near the boundary always has a finite gradient. If the intensity is almost uniform in the interior of Ω and has a sharp decrease near the boundary, then at the boundary the first term on the right-hand side of Eq. (4) is much

larger than the second one, which allows us to write

$$\psi(\theta) \cong -k \frac{\partial_z I(R, \theta)}{\partial_r I(R, \theta)} . \tag{12}$$

Inside Ω the first term on the right-hand side of Eq. (4) is much smaller than the second one, which gives us the expression for f:

$$f(r, \theta) \cong k I_0^{-1} \partial_z I(r, \theta), \qquad r < R . \tag{13}$$

Thus we have a well-defined boundary-value problem [Eqs. (8) and (9)] with the right-hand-side functions f and ψ obtainable from the measurements of optical intensity in two closely spaced planes (we need to measure intensity in two planes in order to calculate the derivative $\partial_z I$). We now proceed with the solution of Eqs. (8) and (9) by the method of orthogonal expansions.

3. TRANSPORT-OF-INTENSITY EQUATION AND ORTHOGONAL POLYNOMIALS

A. Expansion of the Transport-of-Intensity Equation into Orthogonal Polynomials

In this subsection we briefly outline the scheme of the orthogonal expansion method of solution of boundary-value problems for partial differential equations. In the following subsections we will apply it to the TIE using Zernike polynomials, though it is possible to implement this method with any complete set of orthogonal functions. In particular, it may be interesting to consider the eigenfunctions of the Laplacian in the circle. Our choice of Zernike polynomials is motivated by their favorable properties with respect to the description of phase aberrations.[11-14]

Let $\{Z_j\}$ be a complete set of linearly independent functions in domain Ω, so that we can expand the phase ϕ and the wave-front curvature f into a series over Z_j: $\phi = \sum \phi_j Z_j$, $f = \sum f_j Z_j$. Substituting these into the Poisson equation (8) and using the linearity of the Laplace operator, we obtain the system of linear algebraic equations $\sum \Lambda_{ij} \phi_j = f_i$, where the matrix elements Λ_{ij} are the coefficients of the decomposition of $-\Delta Z_j$ over Z_i: $(-\Delta) Z_j = \sum \Lambda_{ij} Z_i$. If the system $\{Z_j\}$ is orthonormal with respect to some scalar product, $\langle Z_i, Z_j \rangle = \delta_{ij}$, where δ_{ij} is the Kronecker symbol, then $\Lambda_{ij} = \langle -\Delta Z_j, Z_i \rangle$. If the matrix $[\Lambda_{ij}]$ is singular, i.e., if it maps some of the Z_j or their linear combinations into zero, then the phase ϕ should be expressed as a sum of two components (projections), $\phi = \phi^{(0)} + \phi^{(1)}$, where $\phi^{(0)}$ is the projection of ϕ onto the subspace $\mathrm{Ker}(-\Delta)$ spanned by all linear combinations of Z_j mapped by the Laplacian into zero, and $\phi^{(1)} = \phi - \phi^{(0)}$ is the projection onto the complementary subspace $[\mathrm{Ker}(-\Delta)]^{\perp}$.

By definition $(-\Delta)\phi^{(0)} = 0$; hence $\phi^{(0)}$ cannot be found from the above system, because $\sum \Lambda_{ij} \phi_j = \sum \Lambda_{ij} \phi_j^{(1)} = f_i$ does not depend on $\phi^{(0)}$. However, if the boundary problem is well posed,[21] $\phi^{(0)}$ can be uniquely found from the boundary conditions. The matrix $[\Lambda_{ij}]$ is always nonsingular (invertible) on the subspace $[\mathrm{Ker}(-\Delta)]^{\perp}$; hence the component $\phi^{(1)}$ can be obtained by the inversion of this matrix: $\phi_j^{(1)} = \sum \Lambda_{ji}^{-1} f_i$. Thus the phase can be reconstructed as a sum of two components, $\phi = \phi^{(0)} + \phi^{(1)}$, with the component $\phi^{(1)}$ obtained from the wave-front curvature and the component $\phi^{(0)}$ determined from boundary conditions.

In practice, we must deal with truncated series, which is equivalent to considering finite-dimensional subspaces spanned by subsets of the whole system $\{Z_j\}$. In what follows, we define such natural subsets and implement the method of orthogonal expansions with the system of Zernike polynomials.

B. Zernike Decomposition of the Laplacian

In this subsection we derive the decomposition of the Laplacian in the spaces $\mathbf{Z}_N$ of Zernike polynomials with radial degree not exceeding some integer N. The main aim is to find the kernel (the polynomials mapped into zero) and the image (the functions into which Zernike polynomials are mapped) of the Laplacian in $\mathbf{Z}_N$. Such an analysis is a necessary preliminary step for the phase reconstruction by the Zernike expansion of the TIE, which we describe in Subsection 3.C. Examples of the subspaces that we introduce in this subsection can be found in Table 1.

We recall the definition of Zernike polynomials using a notation differing from that of Noll[13] only by normalization constants:

$$Z_j(r, \theta) = \begin{cases} c_n^m R_n^m(r)\cos(m\theta) & j \text{ even}, m \neq 0 \\ c_n^m R_n^m(r)\sin(m\theta) & j \text{ odd}, m \neq 0 \\ c_n^0 R_n^0(r) & m = 0 \end{cases}, \tag{14}$$

where $0 \leq r \leq 1$ and $0 \leq \theta \leq 2\pi$,

$$c_n^m = [(2 - \delta_{m0})(n + 1)/\pi]^{1/2} \tag{15}$$

are normalization constants (the factor $\sqrt{\pi}$ in the denominator is the only difference from the notation of Noll[13]),

$$R_n^m(r) = \sum_{s=0}^{(n-m)/2} \gamma_{n,m}^s r^{n-2s} ,$$

$$\gamma_{n,m}^s = \frac{(-1)^s (n-s)!}{s!\,[(n+m)/2 - s]!\,[(n-m)/2 - s]!} \tag{16}$$

are the radial Zernike polynomials, and n and m must be positive integers satisfying $m \leq n$, $n - m$ even. The index m is the azimuthal frequency of a given Zernike polynomial, and n is its radial degree. The number j is a convenient mode-ordering index; it may be verified that in the ordering of Noll[13] each valid pair of indices (m, n), $m \neq 0$, corresponds to the pair of consecutive integer numbers $(j, j + 1)$, where

$$j = j(m, n) = \frac{n(n+1)}{2} + m , \tag{17}$$

whereas each pair $(0, n)$ corresponds to only one number $j + 1$, $j = j(0, n)$ from Eq. (17) (see Table 1). Importantly, this ordering has no gaps; i.e., if we consider the set of all Zernike polynomials with radial degree not exceeding some integer N, their j indices will constitute the set $J_N = \{1, 2, \ldots, j_N\}$ of all consecutive integers from 1 to j_N, where

$$j_N = j(N, N) + 1 = (N + 1)(N + 2)/2 . \tag{18}$$

The choice (15) of normalization constants makes the sys-

Table 1. Spaces of Zernike Polynomials[a]

Radial Degree (n)	Azimuthal Frequency (m): 0	1	2	3	4	5
0	$Z_1 = c$					
1		$Z_2 = c2r\cos\theta$ $Z_3 = c2r\sin\theta$				
2	$Z_4 = c\sqrt{3}(2r^2 - 1)$		$Z_5 = c\sqrt{6}r^2\sin(2\theta)$ $Z_6 = c\sqrt{6}r^2\cos(2\theta)$			
3		$Z_7 = c\sqrt{8}(3r^3 - 2r)\sin\theta$ $Z_8 = c\sqrt{8}(3r^3 - 2r)\cos\theta$		$Z_9 = c\sqrt{8}r^3\sin(3\theta)$ $Z_{10} = c\sqrt{8}r^3\cos(3\theta)$		
4	$Z_{11} = c\sqrt{5}(6r^4 - 6r^2 + 1)$		$Z_{12} = c\sqrt{10}(4r^4 - 3r^2)\cos(2\theta)$ $Z_{13} = c\sqrt{10}(4r^4 - 3r^2)\sin(2\theta)$		$Z_{14} = c\sqrt{10}r^4\cos(4\theta)$ $Z_{15} = c\sqrt{10}r^4\sin(4\theta)$	
5		$Z_{16} = c\sqrt{12}(10r^5 - 12r^3 + 3r)\cos\theta$ $Z_{17} = c\sqrt{12}(10r^5 - 12r^3 + 3r)\sin\theta$		$Z_{18} = c\sqrt{12}(5r^5 - 4r^3)\cos(3\theta)$ $Z_{19} = c\sqrt{12}(5r^5 - 4r^3)\sin(3\theta)$		$Z_{20} = c\sqrt{12}r^5\cos(5\theta)$ $Z_{21} = c\sqrt{12}r^5\sin(5\theta)$

[a]Space $\mathbf{Z}_5$ is spanned by all polynomials from this table, subspace $\mathbf{DZ}_5$ is spanned by all polynomials that are above the diagonal dotted line, subspace $\mathbf{UZ}_5$ is spanned by all polynomials that are below the diagonal dotted line, and subspace $\mathbf{Z}_3 \equiv \mathbf{Z}_{5-2}$ is spanned by all polynomials that are above the horizontal dotted line. Note that $c \equiv 1/\sqrt{\pi}$.

tem of Zernike polynomials orthonormal with respect to the standard scalar product:

$$\langle Z_i, Z_j \rangle = \int_0^{2\pi} \int_0^1 Z_i(r, \theta) Z_j(r, \theta) r \, dr d\theta = \delta_{ij}. \quad (19)$$

Let us introduce the vector space $\mathbf{Z}_N$ of all linear combinations of Zernike polynomials with radial degrees not exceeding an integer N:

$$\mathbf{Z}_N = \left\{ \sum_{j \in J_N} a_j Z_j, \qquad a_j \text{ real numbers} \right\} \quad (20)$$

(see the example for $N = 5$ in Table 1). It follows from Eqs. (18)–(20) that the dimension $\dim(\mathbf{Z}_N)$ of the space $\mathbf{Z}_N$ is equal to j_N: $\dim(\mathbf{Z}_N) = (N + 1)(N + 2)/2$.

We will also need the subspace $\mathbf{DZ}_N$ spanned by all diagonal polynomials from $\mathbf{Z}_N$:

$$\mathbf{DZ}_N = \left\{ \sum a_j Z_j : \qquad Z_j \in \mathbf{Z}_N, \, m = n \right\} \quad (21)$$

(see Table 1). Obviously, $\dim(\mathbf{DZ}_N) = 2N + 1$, as there are two diagonal Zernike polynomials $c_n^n r^n \sin(n\theta)$ and $c_n^n r^n \cos(n\theta)$ for each radial degree $n \neq 0$ and one diagonal polynomial of zero degree, $Z_0 = c_0^0 = \pi^{-1/2}$. Note the important equality

$$\dim(\mathbf{Z}_N) - \dim(\mathbf{DZ}_N) = \dim(\mathbf{Z}_{N-2}), \quad (22)$$

which follows directly from the formulas for the dimensions presented above: $(N + 1)(N + 2)/2 - (2N + 1) = (N - 1)N/2$.

Let us consider the restriction $(-\Delta)_N$ of the Laplacian to the finite-dimensional space $\mathbf{Z}_N$. One can easily verify that the polynomial $(-\Delta)Z_j$ has the same form as that of Z_j [see Eqs. (14)–(16)], with different radial components $\tilde{R}_n^m(r)$ in place of $R_n^m(r)$:

$$\tilde{R}_n^m(r) = \sum_{s=0}^{(n-m)/2} \tilde{\gamma}_{n,m}^s r^{n-2s-2},$$
$$\tilde{\gamma}_{m,n}^s = \gamma_{m,n}^s [m^2 - (n - 2s)^2]. \quad (23)$$

We denote by $\mathrm{Ker}[(-\Delta)_N]$ the kernel of the operator $(-\Delta)_N$, i.e., the set of all functions $\phi \in \mathbf{Z}_N$ mapped by $(-\Delta)_N$ into zero. We state that the kernel of $(-\Delta)_N$ coincides with the space spanned by the diagonal Zernike polynomials:

$$\mathrm{Ker}[(-\Delta)_N] = \mathbf{DZ}_N. \quad (24)$$

In other words, the Laplacian of a linear combination of Zernike polynomials is equal to zero if and only if this linear combination contains only the diagonal polynomials. It is easy to see from Eq. (23) that $\Delta Z_j = 0$ for any diagonal polynomial, which implies that $\mathrm{Ker}[(-\Delta)_N] \supset \mathbf{DZ}_N$. The opposite inclusion, i.e., the fact that any function ϕ from $\mathbf{Z}_N$, such that $\Delta\phi = 0$, can be expressed as a linear combination of diagonal Zernike polynomials, is less obvious and is proved in Appendix A.

Now consider the image space $\mathrm{Im}[(-\Delta)_N]$ of the operator $(-\Delta)_N$, i.e., the set into which $\mathbf{Z}_N$ is mapped by the Laplacian: $\mathrm{Im}[(-\Delta)_N] = (-\Delta)\mathbf{Z}_N$. We state that the image of $(-\Delta)_N$ coincides with the space spanned by the

Zernike polynomials with radial degrees not exceeding $N-2$:

$$\mathrm{Im}[(-\Delta)_N] = \mathbf{Z}_{N-2}. \tag{25}$$

In other words, a function is equal to the Laplacian of a linear combination of Zernike polynomials with radial degrees not exceeding N if and only if it can be expressed as a linear combination of Zernike polynomials with radial degrees not exceeding $N-2$. To prove Eq. (25), we note, first, that according to Eq. (23) the Laplacian decreases the radial degree of any Zernike polynomial by 2. It indicates that $\mathrm{Im}[(-\Delta)_N] \subset \mathbf{Z}_{N-2}$. A rigorous proof of this inclusion is given in Appendix B. Second, the dimension of the image space of any linear operator is always equal to the difference between the dimension of the whole space, where it is defined, and the dimension of its kernel. In the case of $(-\Delta)_N$ we have

$$\begin{aligned}\dim\{\mathrm{Im}[(-\Delta)_N]\} &= \dim(\mathbf{Z}_N) - \dim\{\mathrm{Ker}[(-\Delta)_N]\} \\ &= \dim(\mathbf{Z}_N) - \dim(\mathbf{DZ}_N) = \dim(\mathbf{Z}_{N-2}),\end{aligned} \tag{26}$$

where we used formulas (22) and (24). Thus we see that the image of $(-\Delta)_N$ is a subspace of $\mathbf{Z}_{N-2}$, and its dimension is equal to the dimension of $\mathbf{Z}_{N-2}$. Therefore the vector subspace $\mathrm{Im}[(-\Delta)_N]$ is equal to the whole vector space $\mathbf{Z}_{N-2}$, and Eq. (25) is proved.

Let us denote by $\mathbf{UZ}_N$ the subspace of $\mathbf{Z}_N$ spanned by all nondiagonal polynomials (see Table 1). Because of the orthogonality relations (19), any function ϕ from $\mathbf{Z}_N$ can be uniquely represented as a sum of two orthogonal components:

$$\phi = \phi^{(0)} + \phi^{(1)}, \qquad \phi^{(0)} \in \mathbf{DZ}_N,\ \phi^{(1)} \in \mathbf{UZ}_N. \tag{27}$$

In other words, we have the decomposition of the space $\mathbf{Z}_N$ into an orthogonal sum of two subspaces: $\mathbf{Z}_N = \mathbf{DZ}_N \oplus \mathbf{U}\,\mathbf{Z}_N$, where $\oplus$ denotes the orthogonal sum. Formulas (24) and (25) give us the corresponding decomposition of the Laplacian:

$$(-\Delta)_N \mathbf{Z}_N = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & -\Delta \end{bmatrix} \begin{pmatrix} \mathbf{DZ}_N \\ \mathbf{UZ}_N \end{pmatrix} = \begin{pmatrix} \mathbf{0} \\ \mathbf{Z}_{N-2} \end{pmatrix},$$
$$\text{or } (-\Delta)_N \begin{pmatrix} \phi_{(N)}^{(0)} \\ \phi_{(N)}^{(1)} \end{pmatrix} = \begin{pmatrix} 0 \\ f_{(N-2)} \end{pmatrix}, \tag{28}$$

i.e., operator $(-\Delta)_N$ maps the diagonal subspace $\mathbf{DZ}_N$ into zero, and it maps $\mathbf{UZ}_N$ one to one (bijective) on $\mathbf{Z}_{N-2}$. Hence the restriction of the Laplacian to the subspace of nondiagonal Zernike polynomials is invertible, and corresponding components of the phase can be uniquely found from the wave-front curvature. On the other hand, the diagonal component $\phi_{(N)}^{(0)}$ cannot be obtained from the Poisson equation, and we will need to use the boundary conditions (9) for its determination. The actual algorithm is described in Subsection 3.C.

C. Zernike Decomposition of the Neumann Problem [Eqs. (8) and (9)] and Its Solution

In this subsection we will use the representation (28) of the Laplacian and the Neumann boundary condition (9) to derive an algorithm for the unique reconstruction of the phase ϕ from the wave-front curvature f and the boundary slope ψ.

Suppose that the wave-front curvature f obtained from intensity measurements is approximated by a finite Zernike sum $f_{(N')} \in \mathbf{Z}_{N'}$ with some integer N' (we have explained in Section 2 that the phase solution is stable with respect to small errors in f that may occur as a result of this approximation). We will be looking for a phase ϕ such that $-\Delta\phi = f_{(N')}$. Let $N = N' + 2$. Then $f_{(N')} = f_{(N-2)} \in \mathbf{Z}_{N-2}$, and, according to Eq. (28), we must look for the corresponding phase solution in the space $\mathbf{Z}_N$, i.e., find the coefficients ϕ_j in the representation $\phi_{(N)}(R\rho, \theta) = \sum_{j\in J_N} \phi_j Z_j(\rho, \theta)$, where $\rho = r/R$.

We start with the decomposition of the function $\phi_{(N)} \in \mathbf{Z}_N$ in accordance with Eq. (28):

$$\phi_{(N)} = \phi_{(N)}^{(0)} + \phi_{(N)}^{(1)}, \qquad \phi_{(N)}^{(0)} = \sum_{j\in DJ_N} \phi_j Z_j,$$
$$\phi_{(N)}^{(1)} = \sum_{j\in UJ_N} \phi_j Z_j, \tag{29}$$

where the set DJ_N contains all indices j from $J_N = \{1, 2, \ldots, j_N\}$ corresponding to the diagonal polynomials Z_j and the complementary subset $UJ_N = J_N \backslash DJ_N$ contains all indices corresponding to nondiagonal ones. According to the results of Subsection 3.B, we should be able to retrieve the nondiagonal component $\phi_{(N)}^{(1)}$ from the wave-front curvature $f_{(N-2)}$ using the Poisson equation.

Let us decompose $f_{(N-2)}$ into the Zernike terms,

$$f_{(N-2)}(R\rho, \theta) = \sum_{i\in J_{N-2}} f_i Z_i(\rho, \theta), \tag{30}$$

and use Eq. (28) to obtain

$$\begin{aligned}\sum_{i\in J_{N-2}} f_i Z_i &= f_{(N-2)} = (-\Delta)_N \phi_{(N)}^{(1)} = (-\Delta)_N \sum_{j\in UJ_N} \phi_j Z_j \\ &= R^{-2} \sum_{i\in J_{N-2}} \sum_{j\in UJ_N} \phi_j \Lambda_{ij} Z_i.\end{aligned} \tag{31}$$

Here $\Lambda_{ij} = \langle -\Delta Z_j, Z_i\rangle$ for all $i \in J_{N-2}$, $j \in UJ_N$, with the scalar product $\langle\cdot,\cdot\rangle$ defined in Eq. (19). Note that, on account of the equality (22) and the definition of the set UJ_N, the matrix $[\Lambda_{ij}]$ is square with both dimensions equal to $j_{N-2} = N(N-1)/2$. It is also invertible (nonsingular) because of Eq. (28). As the Zernike polynomials Z_j are linearly independent, all coefficients at Z_i in Eq. (31) with the same indices must be equal, which gives us the following system of linear algebraic equations:

$$\sum_{j\in UJ_N} \Lambda_{ij}\phi_j = R^2 f_i, \qquad i \in J_{N-2}. \tag{32}$$

Solving Eq. (32) for ϕ_j, we obtain

$$\phi_j = R^2 \sum_{i\in J_{N-2}} \Lambda_{ji}^{-1} f_i, \qquad j \in UJ_N, \tag{33}$$

where $[\Lambda_{ji}^{-1}] = [\Lambda_{ij}]^{-1}$ is the inverse matrix [it represents the operator $(-\Delta)_N^{-1}$: $\mathbf{Z}_{N-2} \to \mathbf{UZ}_N$].

Thus we have retrieved the nondiagonal component $\phi_{(N)}^{(1)}$ of the phase using Eq. (33), and it remains for us to find the diagonal component $\phi_{(N)}^{(0)}$. We will use the boundary condition (9) for this purpose. As $\phi_{(N)}$ belongs to the space $\mathbf{Z}_N$, it contains circular harmonics of the orders $m \le N$. Consequently, its normal derivative at the boundary, $\partial_r \phi_{(N)}(R, \theta)$, belongs to the space $\mathbf{F}_N$ spanned

by all circular harmonics with azimuthal frequencies not exceeding N:

$$\mathbf{F}_N = \left\{ \eta(\theta): \quad \eta(\theta) = \eta_0 + \sum_{m=1}^{N} [\eta_m' \sin(m\theta) + \eta_m'' \cos(m\theta)] \right\}. \tag{34}$$

We assume that the boundary wave-front slope ψ obtained from the intensity measurements is approximated by a finite Fourier sum $\psi_{(N)} \in \mathbf{F}_N$:

$$\psi_{(N)}(\theta) = \psi_0 + \sum_{m=1}^{N} [\psi_m' \sin(m\theta) + \psi_m'' \cos(m\theta)]. \tag{35}$$

Then we can write the boundary condition (9) as

$$\partial_r \phi_{(N)}(R, \theta) = \psi_{(N)}(\theta). \tag{36}$$

Obviously, $\partial_r \phi_{(N)}(R, \theta) = \partial_r \phi_{(N)}^{(0)}(R, \theta) + \partial_r \phi_{(N)}^{(1)}(R, \theta)$, and we can calculate the derivative $\partial_r \phi_{(N)}^{(1)}(R, \theta)$, as the $\phi_{(N)}^{(1)}$ component is already found. Let us introduce the new function $\tilde{\psi}_{(N)}(\theta) = \psi_{(N)}(\theta) - \partial_r \phi_{(N)}^{(1)}(R, \theta)$ and decompose it into the circular harmonics:

$$\tilde{\psi}_{(N)}(\theta) = \tilde{\psi}_0 + \sum_{m=1}^{N} [\tilde{\psi}_m' \sin(m\theta) + \tilde{\psi}_m'' \cos(m\theta)]. \tag{37}$$

On the other hand, by its definition $\phi_{(N)}^{(0)}$ is a linear combination of diagonal Zernike polynomials:

$$\phi_{(N)}^{(0)}(R\rho, \theta) = c_0^0 \phi_0 + \sum_{m=1}^{N} c_m^m \rho^m [\phi_m' \sin(m\theta) + \phi_m'' \cos(m\theta)]; \tag{38}$$

hence its radial derivative can be written as

$$\partial_r \phi_{(N)}^{(0)}(R, \theta) = R^{-1} \sum_{m=1}^{N} c_m^m m [\phi_m' \sin(m\theta) + \phi_m'' \cos(m\theta)]. \tag{39}$$

Finally, as $\partial_r \phi_{(N)}^{(0)}(R, \theta) = \tilde{\psi}_{(N)}(\theta)$, we obtain from Eqs. (37) and (39) the following equations for the diagonal coefficients of the phase:

$$\phi_m' = \frac{R\tilde{\psi}_m'}{m c_m^m}, \qquad \phi_m'' = \frac{R\tilde{\psi}_m''}{m c_m^m}, \qquad m = 1, 2, \ldots, N. \tag{40}$$

Formula (40) gives us the values of Zernike coefficients of the diagonal phase component $\phi_{(N)}^{(0)}$, except the zero coefficient ϕ_0. Applying formula (17) with $m = n$, we derive that each integer m in Eq. (40) corresponds to a pair of j indices $(j, j+1) = [m(m+3)/2, m(m+3)/2 + 1]$ of a pair of diagonal polynomials. Hence, when m runs through the sequence $1, 2, \ldots, N$, these pairs $(j, j+1)$ exhaust the whole set DJ_N, except the zero-order Zernike polynomial [note that m starts from unity in Eqs. (39) and (40)].

The last exception has two important consequences. First, the impossibility of determining the coefficient ϕ_0 means that we can reconstruct the phase only up to an arbitrary additive constant. This, however, is a well-known general property of the Neumann problem already discussed in Section 2. Second, boundary conditions (36) cannot be valid [i.e., Eq. (37) cannot be equal to Eq. (39)], unless $\tilde{\psi}_0 = 0$ in Eq. (37). By the conventional formula for the Fourier coefficients we have the equality

$$\tilde{\psi}_0 = \frac{1}{2\pi} \int_0^{2\pi} [\psi_{(N)}(\theta) - \partial_r \phi_{(N)}^{(1)}(R, \theta)] \mathrm{d}\theta = 0. \tag{41}$$

Applying Green's theorem to Eq. (41), we obtain

$$\int_0^{2\pi} \psi_{(N)}(\theta) \mathrm{d}\theta = -R \int_0^{2\pi} \int_0^1 f_{(N-2)}(R\rho, \theta) \rho \mathrm{d}\rho \mathrm{d}\theta. \tag{42}$$

If we substitute the Fourier expansion (35) for $\psi_{(N)}$ and the Zernike expansion (30) for $f_{(N-2)}$ into Eq. (42), we see that the integrals of all the components, except the zero-order ones, are equal to zero because of the orthogonality of the system of circular harmonics and Zernike polynomials. Thus condition (42) is equivalent to

$$\psi_0 = -\frac{R}{2\sqrt{\pi}} f_0, \tag{43}$$

where ψ_0 is the zero-order Fourier coefficient of ψ and f_0 is the zero-order Zernike coefficient of f. We can also obtain Eq. (43) from the energy-conservation law, substituting the Fourier expansion for ψ and the Zernike expansion for f in Eq. (10). Therefore Eq. (43) and hence Eq. (41) are always true.

Thus our reconstruction algorithm allows the unique retrieval by means of formulas (33) and (40) of all phase aberrations with radial degree not exceeding N, except the piston mode. In Subsection 3.D we will derive an explicit matrix formula corresponding to this algorithm, which expresses the reconstructed phase as a function of Zernike coefficients of the wave-front curvature f and Fourier coefficients of the boundary slope ψ.

D. Phase Retrieval Matrix

We consider the operator A_N defined by the Laplacian and the Neumann boundary condition in the space $\mathbf{Z}_N$, $A_N = \{(-\Delta)_N, B_N\}$, with $(B\phi)(\theta) = \partial_r \phi(R, \theta)$, and write the analog of decomposition (28) for it:

$$A_N \mathbf{Z}_N = \begin{bmatrix} B & B \\ \mathbf{0} & -\Delta \end{bmatrix} \begin{pmatrix} \mathbf{DZ}_N \\ \mathbf{UZ}_N \end{pmatrix} = \begin{pmatrix} \mathbf{F}_N \\ \mathbf{Z}_{N-2} \end{pmatrix}, \tag{44}$$

or

$$A_N \begin{pmatrix} \phi_{(N)}^{(0)} \\ \phi_{(N)}^{(1)} \end{pmatrix} = \begin{bmatrix} B_{(N)}^{(0)} & B_{(N)}^{(1)} \\ \mathbf{0} & -\Delta_{(N)} \end{bmatrix} \begin{pmatrix} \phi_{(N)}^{(0)} \\ \phi_{(N)}^{(1)} \end{pmatrix} = \begin{pmatrix} B_{(N)}^{(0)} \phi_{(N)}^{(0)} + B_{(N)}^{(1)} \phi_{(N)}^{(1)} \\ -\Delta \phi_{(N)}^{(1)} \end{pmatrix} = \begin{pmatrix} \psi_{(N)} \\ f_{(N-2)} \end{pmatrix}, \tag{45}$$

where $B_{(N)}^{(0)}$: $\mathbf{DZ}_N \to \mathbf{F}_N$ and $B_{(N)}^{(1)}$: $\mathbf{UZ}_N \to \mathbf{F}_N$ are the restrictions of the boundary operator $(B\phi)(\theta) = \partial_r \phi(R, \theta)$ to the corresponding subspaces. Let us also introduce the spaces $\tilde{\mathbf{F}}_{(N)}$ and $\mathbf{D}\tilde{\mathbf{Z}}_{(N)}$ obtained from $\mathbf{F}_{(N)}$ and $\mathbf{DZ}_{(N)}$, respectively, by the removal of the zero-order (constant) component and define the corresponding operators: $\tilde{B}_{(N)}^{(0)}$: $\mathbf{D}\tilde{\mathbf{Z}}_N \to \tilde{\mathbf{F}}_N$ and $\tilde{B}_{(N)}^{(1)}$: $\mathbf{UZ}_N \to \tilde{\mathbf{F}}_N$. We must remove the zero-order component to make the operator A_N invertible according to Subsection 3.C. Let us rewrite Eq. (45) in

the Zernike and Fourier spaces without constant components:

$$\tilde{A}_N \begin{pmatrix} \tilde{\phi}^{(0)}_{(N)} \\ \phi^{(1)}_{(N)} \end{pmatrix} = \begin{bmatrix} \tilde{B}^{(0)}_{(N)} & \tilde{B}^{(1)}_{(N)} \\ \mathbf{0} & -\Delta_{(N)} \end{bmatrix} \begin{pmatrix} \tilde{\phi}^{(0)}_{(N)} \\ \phi^{(1)}_{(N)} \end{pmatrix} = \begin{pmatrix} \tilde{B}^{(0)}_{(N)}\tilde{\phi}^{(0)}_{(N)} + \tilde{B}^{(1)}_{(N)}\phi^{(1)}_{(N)} \\ -\Delta\phi^{(1)}_{(N)} \end{pmatrix} = \begin{pmatrix} \tilde{\psi}_{(N)} \\ f_{(N-2)} \end{pmatrix}. \quad (46)$$

Now one can easily find the inverse operator $\tilde{A}_N^{-1}$ by matrix algebra:

$$\tilde{A}_N^{-1} = \begin{bmatrix} [\tilde{B}^{(0)}_{(N)}]^{-1} & -[\tilde{B}^{(0)}_{(N)}]^{-1}\tilde{B}^{(1)}_{(N)}[-\Delta_{(N)}]^{-1} \\ \mathbf{0} & [-\Delta_{(N)}]^{-1} \end{bmatrix},$$

$$\tilde{A}_N^{-1}\begin{pmatrix} \tilde{\mathbf{F}}_N \\ \mathbf{Z}_{N-2} \end{pmatrix} = \begin{pmatrix} \mathbf{D}\tilde{\mathbf{Z}}_N \\ \mathbf{U}\mathbf{Z}_N \end{pmatrix}, \quad (47)$$

or

$$\begin{pmatrix} \tilde{\phi}^{(0)}_{(N)} \\ \phi^{(1)}_{(N)} \end{pmatrix} = \tilde{A}_N^{-1}\begin{pmatrix} \tilde{\psi}_{(N)} \\ f_{(N-2)} \end{pmatrix} = \begin{pmatrix} [\tilde{B}^{(0)}_{(N)}]^{-1}\tilde{\psi}_{(N)} - [\tilde{B}^{(0)}_{(N)}]^{-1}\tilde{B}^{(1)}_{(N)}[-\Delta_{(N)}]^{-1}f_{(N-2)} \\ [-\Delta_{(N)}]^{-1}f_{(N-2)} \end{pmatrix}. \quad (48)$$

Formula (48) is an operator representation of the reconstruction algorithm described in Subsection 3.C. Each term in Eq. (48) has already been defined. Operator $[-\Delta_{(N)}]^{-1}$ is defined in Eq. (33). It is represented by the matrix $R^2\Lambda_{ji}^{-1}$ applied to the vector of Zernike coefficients of $f_{(N-2)}$. Operator $[\tilde{B}^{(0)}_{(N)}]^{-1}$ is defined in Eq. (40). Its action on a function $\tilde{\psi}_{(N)}$ from $\tilde{\mathbf{F}}_{(N)}$ is equivalent to the multiplication of its Fourier coefficients $\tilde{\psi}'_m$, $\tilde{\psi}''_m$ by the constants R/mc_m^m. Operator $\tilde{B}^{(1)}_{(N)}$ is the restriction of the boundary operator $(B\phi)(\theta) = \partial_r\phi(R, \theta)$, $\tilde{B}^{(1)}_{(N)}$: $\mathbf{UZ}_N \rightarrow \tilde{\mathbf{F}}_N$. The inversion (48) admits a generalization onto the infinite-dimensional case,[24] but we will not need it in this paper.

Thus we have obtained the explicit formula (48) for the operator $\tilde{A}_N^{-1}$, which maps each pair $\{\psi_{(N)}, f_{(N-2)}\}$ of the wave-front curvature $f_{(N-2)} \in \mathbf{Z}_{N-2}$ and boundary slope $\psi_{(N)} \in \mathbf{F}_N$ onto the unique (up to an additive constant) phase solution $\tilde{\phi}_{(N)} = \tilde{\phi}^{(0)}_{(N)} + \phi^{(1)}_{(N)} \in \tilde{\mathbf{Z}}_N$ such that $-\Delta\tilde{\phi}_{(N)} = f_{(N-2)}$ and $\partial_r\tilde{\phi}_{(N)}(R, \theta) = \psi_{(N)}(\theta)$. Formula (48) is the central result of the present paper.

Now we will give the analytical form for the matrix formulas (46)–(48) convenient for applications and present a numerical example for $N = 4$. Let us start from the matrix $\tilde{A}_N$:

$$\tilde{A}_N = \begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{bmatrix} = \begin{bmatrix} R^{-1}B^{(0)}_{ij} & R^{-1}B^{(1)}_{ij} \\ 0 & R^{-2}\Lambda_{ij} \end{bmatrix}. \quad (49)$$

In Eq. (49) $A_{11} = R^{-1}B^{(0)}_{ij}$ is a square $2N \times 2N$ matrix, and $A_{12} = R^{-1}B^{(1)}_{ij}$ is a rectangular $2N \times (N-1)N/2$ matrix with elements

$$R^{-1}B_{ij} = \langle BZ_j, F_i\rangle_\Gamma = R^{-1}\int_0^{2\pi}(\partial_\rho Z_j)(1, \theta)F_i(\theta)\mathrm{d}\theta, \quad (50)$$

where

$$F_i = \begin{cases} c\cos\left(\frac{i}{2}\theta\right) & i \text{ even} \\ c\sin\left(\frac{i+1}{2}\theta\right) & i \text{ odd} \end{cases}, \qquad c = 1/\sqrt{\pi}, \quad (51)$$

where $i = 1, 2, \ldots, 2N$, $j \in DJ_N\backslash\{1\}$ for $B^{(0)}_{ij}$, and $j \in UJ_N$ for $B^{(1)}_{ij}$. Using Eq. (50) and the definition (14)–(16) of the Zernike polynomials, we obtain

$$R^{-1}B^{(0)}_{ij} = R^{-1}mc_m^m\delta_{ik}, \quad i = 1, 2, \ldots, 2N, \quad j \in DJ_N\backslash\{1\}, \quad m = m(j), \quad (52)$$

$$R^{-1}B^{(1)}_{ij} = R^{-1}\sigma_j\delta_{ik}, \qquad i = 1, 2, \ldots, 2N, \; j \in UJ_N. \quad (53)$$

$$k = \begin{cases} 2m(j) - 1 & j \text{ odd} \\ 2m(j) & j \text{ even} \end{cases}, \quad (54)$$

where $\sigma_j = c_n^m \sum_{s=0}^{(n-m)/2} \gamma^s_{n,m}(n - 2s)$ [see Eqs. (15) and (16) for the definition of the coefficients c_n^m and $\gamma^s_{n,m}$], δ_{ik} is the Kronecker symbol, and $m(j)$ is the azimuthal frequency of Z_j. The block $A_{22} = R^{-2}\Lambda_{ij}$ is the $(N-1)N/2 \times (N-1)N/2$ square matrix with elements

$$R^{-2}\Lambda_{ij} = \langle -\Delta Z_j, Z_i\rangle = \int_0^{2\pi}\int_0^R (-\Delta)Z_j(r/R, \theta)Z_i(r/R, \theta)r\,\mathrm{d}r\mathrm{d}\theta, \quad (55)$$

where $i \in J_{N-2}$ and $j \in UJ_N$. The following formula is derived from the definition (14)–(16) of Zernike polynomials and formula (23) for the Laplacian of Zernike polynomials:

$$\Lambda_{ij} = \delta_{mm'}2[(n+1)(n'+1)]^{1/2}\sum_{s'=0}^{(n'-m)/2}\sum_{s=0}^{(n-m)/2}\gamma^{s'}_{n',m}\gamma^s_{n,m} \times \frac{m^2 - (n-2s)^2}{n' + n - 2s' - 2s}, \quad (56)$$

where $i \in J_{N-2}$, $j \in UJ_N$, $j = j(m, n)$, and $i = j(m', n')$, as in Eq. (17), and the coefficients $\gamma^s_{n,m}$ are defined in Eq. (16). Note that although formulas (52), (53), and (56) give explicit analytical expressions for the elements of the matrix $\tilde{A}_{(N)}$, it is sometimes more convenient to use formulas (50) and (55) for the practical calculations. In Table 2 we present an example of matrix $\tilde{A}_{(N)}$ for $N = 4$.

The inverse matrix $\tilde{A}_{(N)}^{-1}$ consists of the blocks

$$\tilde{A}_N^{-1} = \begin{bmatrix} A_{11}^{-1} & A_{12}^{-1} \\ 0 & A_{22}^{-1} \end{bmatrix} = \begin{bmatrix} R[B^{(0)}_{ij}]^{-1} & -R^2[B^{(0)}_{ij}]^{-1}B^{(1)}_{ij}\Lambda_{ji}^{-1} \\ 0 & R^2\Lambda_{ji}^{-1} \end{bmatrix}. \quad (57)$$

In Eq. (57) A_{11}^{-1} is a square $2N \times 2N$ matrix, A_{12}^{-1} is a rectangular $2N \times (N-1)N/2$ matrix, and A_{22}^{-1} is an $(N-1)N/2 \times (N-1)N/2$ square matrix. In order to obtain the elements of $\tilde{A}_{(N)}^{-1}$, one needs to find only the inverse matrices $\Lambda_{ji}^{-1} = (\Lambda_{ij})^{-1}$ and $[B^{(0)}_{ij}]^{-1}$. The elements of these inverse matrices can be easily calculated:

Table 2. Matrix $\tilde{A}_{(4)}$ for $R = 1$[a]

	j													
i	2	3	5	6	9	10	14	15	4	7	8	11	12	13
1	0	2	0	0	0	0	0	0	0	$14\sqrt{2}$	0	0	0	0
2	2	0	0	0	0	0	0	0	0	0	$14\sqrt{2}$	0	0	0
3	0	0	$2\sqrt{6}$	0	0	0	0	0	0	0	0	0	0	$10\sqrt{10}$
4	0	0	0	$2\sqrt{6}$	0	0	0	0	0	0	0	0	$10\sqrt{10}$	0
5	0	0	0	0	$6\sqrt{2}$	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0	$6\sqrt{2}$	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	$4\sqrt{10}$	0	0	0	0	0	0
8	0	0	0	0	0	0	$4\sqrt{10}$	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	$-8\sqrt{3}$	0	0	$-24\sqrt{5}$	0	0
2	0	0	0	0	0	0	0	0	0	0	$-24\sqrt{2}$	0	0	0
3	0	0	0	0	0	0	0	0	0	$-24\sqrt{2}$	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0	0	$-16\sqrt{15}$	0	0
5	0	0	0	0	0	0	0	0	0	0	0	0	0	$-16\sqrt{15}$
6	0	0	0	0	0	0	0	0	0	0	0	0	$-16\sqrt{15}$	0

[a]The rules separate the blocks A_{pq}. The indexing system is described in the text [formulas (49) and (52)–(56)].

Table 3. Phase Retrieval Matrix $\tilde{A}_{(4)}^{-1}$ for $R = 1$[a]

	i													
j	1	2	3	4	5	6	7	8	1	2	3	4	5	6
2	0	1/2	0	0	0	0	0	0	0	7/24	0	0	0	0
3	1/2	0	0	0	0	0	0	0	0	0	7/24	0	0	0
5	0	0	$1/(2\sqrt{6})$	0	0	0	0	0	0	0	0	0	5/48	0
6	0	0	0	$1/(2\sqrt{6})$	0	0	0	0	0	0	0	0	0	5/48
9	0	0	0	0	$1/(6\sqrt{2})$	0	0	0	0	0	0	0	0	0
10	0	0	0	0	0	$1/(6\sqrt{2})$	0	0	0	0	0	0	0	0
14	0	0	0	0	0	0	0	$1/(4\sqrt{10})$	0	0	0	0	0	0
15	0	0	0	0	0	0	$1/(4\sqrt{10})$	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	$-1/(8\sqrt{3})$	0	0	1/16	0	0
7	0	0	0	0	0	0	0	0	0	0	$-1/(24\sqrt{2})$	0	0	0
8	0	0	0	0	0	0	0	0	0	$-1/(24\sqrt{2})$	0	0	0	0
11	0	0	0	0	0	0	0	0	0	0	0	$-1/(16\sqrt{15})$	0	0
12	0	0	0	0	0	0	0	0	0	0	0	0	0	$-1/(16\sqrt{15})$
13	0	0	0	0	0	0	0	0	0	0	0	0	$-1/(16\sqrt{15})$	0

[a]The rules separate the blocks A_{pq}^{-1}. The first 14 nonconstant Zernike aberrations ($j = 2, 3, \ldots, 15$) of a phase can be retrieved by the application of this matrix to the vector $\{\psi_{(4)}, f_{(2)}\}$ of the Fourier coefficients of the boundary slope and the Zernike coefficients of the curvature of the wave front. The indexing system is described in the text [formulas (57)–(60)].

$$(A_{11}^{-1})_{ji} = R\delta_{ik}/mc_m^m,$$

$$i = 1, 2, \ldots, 2N,\ j \in DJ_N\backslash\{1\},\ m = m(j), \quad (58)$$

$$(A_{22}^{-1})_{ji} = R^2(-1)^{i+j}M_{ij}/\det\Lambda, \qquad i \in J_{N-2},\ j \in UJ_N, \quad (59)$$

where $\det\Lambda$ is the determinant of the $(N-1)N/2 \times (N-1)N/2$ matrix Λ_{ij} with the elements defined in Eq. (56) and M_{ij} is the determinant of the matrix obtained from Λ_{ij} by removing the ith row and the jth column. Thus we defined all the elements of the diagonal blocks A_{11}^{-1} and A_{22}^{-1}. The elements of the nondiagonal block A_{12}^{-1} can now be obtained according to the rules of matrix multiplication:

$$(A_{12}^{-1})_{ji} = -R^2\sum_k\sum_l (A_{11}^{-1})_{jk}B_{kl}^{(1)}(A_{22}^{-1})_{li},$$

$$j \in DJ_N\backslash\{1\},\ i \in J_{N-2}. \quad (60)$$

Formulas (58)–(60) give explicit analytical expressions for the elements of the matrix $\tilde{A}_N^{-1}$; however, in practice, it may be easier to obtain the whole matrix $\tilde{A}_{(N)}^{-1}$ by the matrix inversion of $\tilde{A}_{(N)}$.

The phase $\phi_{(N)} \in \mathbf{Z}_{(N)}$ is retrieved by the direct multiplication of this matrix $\tilde{A}_N^{-1}$ and the vector $g_{(N)} = \{\psi_{(N)}, f_{(N-2)}\}$ of the Fourier coefficients of boundary slopes $\psi_{(N)} \in \mathbf{F}_N$ and the Zernike coefficients of the wave-front curvature $f_{(N-2)} \in \mathbf{Z}_{N-2}$:

$$\begin{pmatrix} \tilde{\phi}_{(N)}^{(0)} \\ \phi_{(N)}^{(1)} \end{pmatrix} = \tilde{A}_N^{-1}\begin{pmatrix} \psi_{(N)} \\ f_{(N-2)} \end{pmatrix} = \begin{bmatrix} A_{11}^{-1} & A_{12}^{-1} \\ 0 & A_{22}^{-1} \end{bmatrix}\begin{pmatrix} \psi_{(N)} \\ f_{(N-2)} \end{pmatrix}. \quad (61)$$

As can be seen from Eqs. (57)–(60), the phase retrieval matrix $\tilde{A}_{(N)}^{-1}$ does not depend on experimental data and can be calculated for a given N once and for all with arbitrary precision. In fact, using algebraic, rather than numerical, calculations with the help of such mathematical software as, for example, MATHEMATICA, one can calculate the matrix $\tilde{A}_{(N)}^{-1}$ with absolute precision. We performed this algebraic calculation for $N = 4$, with the result presented in Table 3.

It is worth noting that the matrix in Table 3 is very sparse and that the values of its nonzero elements are decreasing as the matrix indices increase. This indicates that the retrieval of phase with this matrix should be quite stable. We still expect larger errors in the retrieval of the diagonal (with $m = n$) Zernike aberrations compared with those arising from the nondiagonal terms (with $m \neq n$), because of the necessity to distinguish the contribution of the wave-front boundary slopes from that of the wave-front curvature. As can be seen from Eq. (61), if the uncertainties in $\psi_{(N)}$ are greater than the uncertainties in $f_{(N-2)}$, then the uncertainty in the reconstruction of the diagonal phase component $\tilde{\phi}_{(N)}^{(0)}$ will also be greater than that in the nondiagonal component $\phi_{(N)}^{(1)}$, as the latter depends only on the wave-front curvature $f_{(N-2)}$. A full discussion of the stability phase retrieval with this approach is beyond the scope of the present paper, and a thorough study of stability is currently under way.

4. SUMMARY

In this paper we have examined the solution of the transport-of-intensity equation (TIE) from the perspective of a Zernike polynomial decomposition. We have found a direct matrix relationship between a Zernike polynomial decomposition of the intensity derivative and the Zernike decomposition of the phase. This relationship gives a precise description of the influence of individual Zernike aberrations of the phase on the evolution of the intensity distribution in the wave front near a uniformly illuminated circular aperture, making phase retrieval extremely straightforward. The derived operator solution (48) to the Neumann boundary problem for the TIE and its matrix representation (61) give a unique phase $\phi_{(N)} \in \mathbf{Z}_{(N)}$ for any given wave-front curvature inside the aperture $f_{(N-2)} \in \mathbf{Z}_{(N-2)}$ and the boundary slope $\psi_{(N)} \in \mathbf{F}_{(N)}$. In practice, $\phi_{(N)}$ is obtained as a sum of two orthogonal components $\phi_{(N)}^{(0)}$ and $\phi_{(N)}^{(1)}$, the first containing all diagonal Zernike aberrations of the phase and the second containing all the nondiagonal ones. The component $\phi_{(N)}^{(1)}$ depends only on the wave-front curvature inside the aperture. The diagonal component $\phi_{(N)}^{(0)}$ depends on the wave-front slope at the boundary as well as on the boundary values of $\phi_{(N)}^{(1)}$. We have also derived explicit analytical expressions [Eqs. (57)–(60)] for the phase retrieval matrix $\tilde{A}_N^{-1}$ corresponding to the operator solution (48) and presented a numerical example of such a matrix for $N = 4$. The phase $\phi_{(N)} \in \mathbf{Z}_{(N)}$ can be retrieved by the direct multiplication of the matrix $\tilde{A}_N^{-1}$ with the vector $g_{(N)} = \{\tilde{\psi}_{(N)}, f_{(N-2)}\}$ of the Fourier coefficients of boundary slopes and the Zernike coefficients of the wave-front curvature. The matrix $\tilde{A}_N^{-1}$ does not depend on the experimental data and so needs to be calculated only once for a given N. Thus there is the prospect of a very rapid and precise phase retrieval algorithm with this approach. We are planning to present results of computer simulations with this algorithm in a subsequent paper.

APPENDIX A

Here we prove that the kernel of the Laplacian in the Zernike space $\mathbf{Z}_N$ coincides with the subspace $\mathbf{DZ}_N$ spanned by the diagonal Zernike polynomials, i.e., $\mathrm{Ker}[(-\Delta)_N] = \mathbf{DZ}_N$. It was shown in Subsection 3.B that $\Delta Z_j = 0$ for any diagonal polynomial, so $\mathrm{Ker}[(-\Delta)_N] \supset \mathbf{DZ}_N$. Hence it is sufficient to prove the opposite inclusion,

$$\mathrm{Ker}[(-\Delta)_N] \subset \mathbf{DZ}_N , \tag{A1}$$

i.e., that any function g from $\mathbf{Z}_N$, such that $\Delta\phi = 0$, can be expressed as a linear combination of diagonal Zernike polynomials. Let us consider an arbitrary function ϕ from $\mathbf{Z}_N$:

$$\phi = \sum_{j \in J_N} \phi_j Z_j = \sum_n \sum_m r^n[\phi'_{nm} \sin(m\theta) + \phi''_{nm} \cos(m\theta)], \tag{A2}$$

where the first sum is over $n \leq N$ and the second is over such m that $n - m$ is positive and even. Then

$$-\Delta\phi = \sum_n \sum_m (m^2 - n^2) r^{n-2}[\phi'_{nm} \sin(m\theta) + \phi''_{nm} \cos(m\theta)]. \tag{A3}$$

All the monomials $r^{n-2}\sin(m\theta)$ and $r^{n-2}\cos(m\theta)$ in Eq. (A3) are linearly independent, so $\Delta\phi = 0$ if and only if all the coefficients in Eq. (A3) are equal to zero:

$$(m^2 - n^2)\phi'_{nm} = 0, \qquad (m^2 - n^2)\phi''_{nm} = 0 . \tag{A4}$$

But Eqs. (A4) imply that all coefficients ϕ'_{nm} and ϕ''_{nm} with $n \neq m$ must be equal to zero, so in the expansion (A2), only diagonal coefficients ϕ'_{mm} and ϕ''_{mm} may be nonzero, i.e., the function ϕ belongs to the diagonal subspace $\mathbf{DZ}_N$ and relation (A1) is proved.

APPENDIX B

Here we prove that the Laplacian maps the Zernike space $\mathbf{Z}_N$ into the space $\mathbf{Z}_{N-2}$, i.e.,

$$\mathrm{Im}[(-\Delta)_N] \subset \mathbf{Z}_{N-2} . \tag{B1}$$

In other words, we must prove that for any function ϕ from $\mathbf{Z}_N$ its Laplacian $\Delta\phi$ can be represented as a linear combination of some Zernike polynomials with radial degrees not exceeding $N - 2$. As $\Delta\phi$ can be expressed by formula (A3), it is sufficient to prove that each monomial $r^{n-2}\sin(m\theta)$ and $r^{n-2}\cos(m\theta)$ in Eq. (A3) belongs to $\mathbf{Z}_{N-2}$. Importantly, $m \leq (n - 2)$ for all terms in Eq. (A3), because the terms with $m = n$ from Eq. (A2) are mapped by the Laplacian into zero. We consider only monomials with sin (the proof for the monomials with cos is identical). Thus we are going to prove that any monomial $r^l \sin(m\theta)$, with $l \leq n - 2$, $l - m$ nonnegative and even, belongs to $\mathbf{Z}_{N-2}$. It is then sufficient to prove that for any $m \leq N - 2$ the radial monomial r^l, $l = m + 2k$, $l \leq N - 2$, can be represented as a linear combination

$$r^l = \sum_{n \leq l} \sigma_n R_n^m(r) \tag{B2}$$

of radial Zernike polynomials with fixed m. We will prove it by induction by k. For $k = 0$ we have $l = m$ and Eq. (B2) is obvious, as $r^m = R_m^m$ [see Eq. (16)]. Let us assume that we have proved Eq. (B2) for $l = m + 2(k - 1)$ and then prove it for $l = m + 2k$. Using the

definition (16) of Zernike polynomials, we can write

$$r^{m+2k} = \left[R^{m}_{m+2k} - \sum_{s=1}^{k} \gamma^{s}_{m+2k,m} r^{m+2(k-s)} \right] \Big/ \gamma^{0}_{m+2k,m} . \quad \text{(B3)}$$

But by the induction assumption all monomials under the summation sign in Eq. (B3) can be represented in the form (B2) [as their indices do not exceed $m + 2(k - 1)$], and R^{m}_{m+2k} obviously has the form (B2) with $m + 2k = l \leq N - 2$. Therefore r^{m+2k} can be represented as the linear combination (B2), and hence Eq. (B1) is proved.

ACKNOWLEDGMENTS

This research was supported by the Australian Research Council. The authors acknowledge helpful discussions with T. J. Davis, A. W. Stevenson, and D. Paganin.

REFERENCES

1. R. P. Millane, "Phase retrieval in crystallography and optics," J. Opt. Soc. Am. A **7**, 394–411 (1990).
2. M. Nieto-Vesperinas, *Scattering and Diffraction in Physical Optics* (Wiley, New York, 1991), Sect. 9.10.
3. M. R. Teague, "Irradiance moments: their propagation and use for unique retrieval of phase," J. Opt. Soc. Am. **72**, 1199–1209 (1982).
4. M. R. Teague, "Deterministic phase retrieval: a Green's function solution," J. Opt. Soc. Am. **73**, 1434–1441 (1983).
5. F. Roddier, "Curvature sensing and compensation: a new concept in adaptive optics," Appl. Opt. **27**, 1223–1225 (1988).
6. F. Roddier, "Wavefront sensing and the irradiance transport equation," Appl. Opt. **29**, 1402–1403 (1990).
7. N. Streibl, "Phase imaging by the transport equation of intensity," Opt. Commun. **49**, 6–10 (1984).
8. J. Liang, B. Grimm, S. Goelz, and J. F. Bille, "Objective measurement of wave aberrations of the human eye with the use of a Hartmann–Shack wave-front sensor," J. Opt. Soc. Am. A **11**, 1949–1957 (1994).
9. J. Susini, G. Marot, L. Zhang, R. Ravelet, and P. Jagourel, "Conceptual design of an adaptive x-ray mirror prototype for the ESRF," Rev. Sci. Instrum. **63**, 489–492 (1992).
10. C. Roddier and F. Roddier, "Wave-front reconstruction from defocused images and the testing of ground-based optical telescopes," J. Opt. Soc. Am. A **10**, 2277–2287 (1993).
11. M. Born and E. Wolf, *Principles of Optics*, 6th ed. (Pergamon, Oxford, 1980), Sect. 9.2.
12. S. N. Bezdid'ko, "The use of Zernike polynomials in optics," Sov. J. Opt. Technol. **41**, 425–429 (1974).
13. R. J. Noll, "Zernike polynomials and atmospheric turbulence," J. Opt. Soc. Am. **66**, 207–211 (1976).
14. W. J. Tango, "The circle polynomials of Zernike and their application in optics," Appl. Phys. **13**, 327–332 (1977).
15. J. Y. Wang and D. E. Silva, "Wave front interpretation with Zernike polynomials," Applied Opt. **19**, 1510–1518 (1980).
16. V. N. Mahajan, "Zernike circle polynomials and optical aberrations of systems with circular pupils," Eng. Lab. Notes **17**, S21–S24 (1994).
17. C. Schwartz, E. Ribak, and S. G. Lipson, "Bimorph adaptive mirrors and curvature sensing," J. Opt. Soc. Am. A **11**, 895–902 (1994).
18. V. N. Mahajan, "Zernike annular polynomials for imaging systems with annular pupils," J. Opt. Soc. Am. **71**, 75–85 (1981); **71**, 1408 (1981); A **1**, 685 (1984).
19. W. Swantner and W. W. Chow, "Gram–Schmidt orthonormalization of Zernike polynomials for general aperture shapes," Appl. Opt. **33**, 1832–1837 (1994).
20. P. Hickson, "Wave-front curvature sensing from a single defocused image," J. Opt. Soc. Am. A **11**, 1667–1673 (1994).
21. S. G. Mikhlin, *Mathematical Physics, An Advanced Course* (North-Holland, Amsterdam, 1970).
22. T. E. Gureyev, A. Roberts, and K. A. Nugent, "Partially coherent fields, the transport-of-intensity equation, and phase uniqueness," J. Opt. Soc. Am. A **12**, 1942–1946 (1995).
23. F. Gori, M. Santasiero, and G. Guattari, "Coherence and the spatial distribution of intensity," J. Opt. Soc. Am. A **10**, 673–679 (1993).
24. S. Rempel and B.-W. Schulze, *Index Theory of Elliptic Boundary Value Problems* (Akademie-Verlag, Berlin, 1982).